Berücksichtigung von systematischen Modellabweichungen in der Kanalnetzsimulation

Dario Del Giudice[1], Mark Honti[1], Andreas Scheidegger[1], Carlo Albert[1], Peter Reichert[1], and Jörg Rieckermann[1*]

[1]Eidgen. Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz (Eawag), Überlandstrasse 133, CH-8600 Dübendorf
[2]ETH Zürich, Institut für Umweltsystemwissenschaften, Wolfgang-Pauli-Str. 15, CH-8093 Zürich

*Email des korrespondierenden Autors: joerg.rieckermann@eawag.ch

Kurzfassung Hydrodynamische Kanalnetzmodelle sind für die Entwässerungsplanung unersetzlich. Obwohl sie den komplexen Niederschlags-Abfluss Prozess niemals exakt abbilden können, liefern sie oft akzeptable Ergebnisse. Es ist jedoch schwierig, plausible Schätzungen der Unsicherheit von Kanalnetzberechnungen zu erhalten, da die gängigen statistischen Verfahren auf sehr starken Annahmen über die Fehlerverteilung beruhen, wie z.B. unabhängige, und identisch verteilte Zufallseffekte, welche praktisch nie erfüllt sind. In diesem Artikel stellen wir deshalb eine Methode vor, um i) zuverlässigere Schätzungen der Unsicherheit von Kanalnetzsimulationen zu berechnen und ii) die Vorhersageunsicherheit in einzelne Komponenten zu trennen. Zu diesem Zweck werden neben dem deterministischen Modell und zufälligen Messfehlern auch systematische Modellabweichungen berücksichtigt. Diese systematischen Modellabweichungen werden durch einen autoregressiven Zufallsprozess beschrieben und erfassen die Effekte von Unsicherheiten in den Eingangsdaten und der Modellstruktur. Zur Kalibration wird Bayessche Parameterschätzung verwendet. Wir zeigen an einem Fallbeispiel, dass die Berücksichtigung der Modellabweichung die Zuverlässigkeit der Berechnungsergebnisse stark verbessert. Ausserdem wird deutlich, dass systematische Modellabweichungen die Vorhersageunsicherheit dominieren und weit grösser sind, als Unsicherheit in den Modellparametern oder Messfehler. Abgesehen von den langen Rechenzeiten ergeben sich gewisse Limitierungen aus der subjektiven Formulierung des Vorwissens über die Parameter des Fehlermodells. Die statistische Berücksichtigung von systematischen Modellabweichungen liefert jedoch eine zuverlässige Schätzungen der Unsicherheit als herkömmliche Methoden und scheint daher für Kanalnetzsimulationen sehr interessant.

Schlagwörter: Kanalnetzsimulation, Unsicherheit, systematische Modellabweichungen, Bayessche Parameterschätzung

1 EINLEITUNG

In der Stadtentwässerung sind Modellprognosen des Abflussverhaltens von Kanalisationen von größter Bedeutung für den Überflutungsschutz. Leider sind Modellprognosen immer mit Fehlern behaftet und die Rechenergebnisse daher ungenau. Die relevanten Fehlerquellen lassen sich typischerweise unterteilen in: i) unsichere Eingangsdaten, ii) mangelhaftes Wissen über die optimalen Modellparameter, iii) systematische Abweichungen, z.B. durch eine vereinfachte Modellstruktur und iv) zufällige Messfehler, z.B. in Abflussmessungen. In den letzten Jahren ist die Unsicherheit von Modellvorhersagen stärker in den Fokus von Wissenschaft und Praxis gerückt. Häufig werden dazu klassische statistische Methoden eingesetzt. Diese basieren jedoch auf starken Annahmen: die Abweichungen von Modellprognosen und Messdaten seien unkorreliert und identisch verteilt. Allerdings werden aufgrund der im Vergleich zur Realität sehr vereinfachten Modellstruktur, und vergleichsweise guter Abflussmessungen, diese Annahmen bei Kanalnetzsimulationen praktisch immer verletzt (Dotto et al., 2012)

In diesem Beitrag schlagen wir, basierend auf der Arbeiten von Reichert und Schuwirth (2012), eine neue Methode vor, mit der sich alle oben beschriebenen Fehlerquellen statistisch berücksichtigen lassen. So lässt sich nicht nur die gesamte Unsicherheit von Modellprognosen verlässlich berechnen, sondern auch eine Aussage über den Einfluss der einzelnen Fehlertypen machen. Die Grundidee der Methode besteht darin, realistischere Annahmen über den zugrundeliegenden Fehlerprozess, resp. die Residuen, zu machen. Das heisst, dass autokorrelierte Fehler explizit berücksichtigt werden. Insbesondere schlagen wir einen strukturierten Ansatz vor, um die beste Beschreibung der Modellabweichung für eine bestimmtes

Entwässerungssystem und deterministisches Modell zu erarbeiten. Wir testen unseren Ansatz an einem realen Regenwasser-System in Prag, Tschechische Republik, und diskutieren die Ergebnisse, die mit verschiedenen mathematischen Formulierungen der Modellabweichung erzielt werden. Eine ausführliche Beschreibung ist in Del Giudice et al. (2013) zu finden.

2 METHODEN

2.1 Beschreibung der Modellabweichung durch einen zusätzlichen Fehlerterm

Mathematisch werden die Messdaten Y_0 häufig durch die Summe des deterministischen Modellergebnisse und einer Zufallsvariable (E) beschrieben. Traditionell werden die Modellresultate als „wahre" Grössen und die Zufallsvariable als zufällige Messfehler interpretiert, welcher durch ein statistisches Fehlermodell mit unabhängigen und gleichverteilten Fehlern beschrieben werden. Da Kanalnetzmodelle nie exakt das "wahre" Verhalten eines Entwässerungssystems abbilden können, schlagen wir vor, zusätzlich zu den Messfehlern ein autoregressives Fehlermodell zu verwenden. Dieses autoregressive Fehlermodell beschreibt dann die Modellabweichungen mit einem zusätzlichen additiven (B_M):

$$\widetilde{Y}_0(x,\theta,\psi) = \widetilde{y_M}(x,\theta) + B_M(x,\psi) + E(\psi) \qquad \text{Gl. (1)}$$

Dabei erfasst B_M alle systematischen Abweichungen, die z.B. durch schlechte Eingangsdaten, eine zu einfache Modellstruktur, systematische Messfehler, oder deren Kombinationen verursacht werden. In seiner einfachsten Form wird die Modellabweichung als autokorrelierter stationärer zufälliger Prozess beschrieben (Reichert and Schuwirth 2012). Manchmal können aber auch komplexere Formulierungen sinnvoll sein, beispielsweise eine Abhängigkeit von den Niederschlagsdaten (Honti et al., 2013). Für praktische Anwendungen ist die Wahl einer adäquaten Struktur für den autokorrelierten Prozess nicht trivial. Einerseits fehlen Beispiele aus der Stadtentwässerung, andererseits resultieren die Abweichungen aus dem komplexen Zusammenspiel zwischen dem realen Verhalten des Entwässerungsnetzes, dem Modell und der Güte der Messdaten. Im Folgenden werden vier unterschiedliche Formulierungen für die Varianz des stochastischen Prozesses untersucht: i) konstant, ii) abhängig vom Niederschlag, iii) abhängig vom Durchfluss, iv) niederschlags- und durchflussabhängig. Die konstante Abweichung wird über einen einfachen Ornstein-Uhlenbeck (OU) Prozess modelliert. Die Niederschlagsabhängigkeit nutzt einen modifizierten OU-Prozess, dessen Varianz durch Regen vergrössert wird. Die Durchflussabhängigkeit wird durch Transformation der Messdaten und Modellergebnisse erreicht. Basierend auf diesen Annahmen kann dann die Likelihood-Funktion formuliert werden, welche zur Modellkalibration und zur Berechnung der Vorhersageunsicherheit benötigt wird.

Die Likelihood-Funktion $f(y_0|\xi, x)$ wird für ein deterministisches Modell, M, und einen zufälligen Term formuliert. $f(y_0|\theta,\psi,x)$ beschreibt die gemeinsame Wahrscheinlichkeitsdichte der beobachteten Messdaten, y_o, gegeben die Parameter des deterministischen Modells (θ) und des Fehlermodells (ξ) und Eingangsgrössen des Systems, x, wie beispielsweise den beobachteten Niederschlag. Für alle untersuchten Formulierungen der Modellabweichung ist die Likelihood-Funktion die Dichtefunktion einer multivariate Normalverteilung mit der Kovarianzmatrix $\Sigma(\psi, x)$:

$$f(y_0|\theta,\psi,x) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{det(\Sigma(\psi,x))}} exp\left(-\frac{1}{2}[\widetilde{y_0} - \widetilde{y_M}(\theta,x)]^T \Sigma(\psi,x)^{-1} [\widetilde{y_0} - \widetilde{y_M}(\theta,x)]\right) \prod_{i=1}^{n} \frac{dg}{dy}(y_{0,i},\psi) \qquad \text{Gl. (2)}$$

wobei n die Anzahl der Beobachtungen sind, z.B. für eine Abflussmessung, $y_M(\theta, x)$ sind die deterministischen Modellresultate und $g(.)$ ist eine Transformationsfunktion. Transformierte Grössen sind mit einer Tilde gekennzeichnet, d.h. $\tilde{y} = g(y)$. Die je nach Annahme wird Kovarianzmatrix angepasst. Zudem sind Transformation eine gängige Technik, um die zugrundeliegenden statistischen Annahmen zu ändern. Wenn dies nicht genügt, sind gute Strategien (i) die Verbesserung der Modellstruktur, (ii) die Modifizierung der Verteilungsannahmen, zum Beispiel durch eine Transformation oder (iii) die statistische Beschreibung der Modellabweichungen. Im Folgenden schlagen wir eine Kombination von (ii) und (iii) vor.

2.2 Statistische Formulierung der systematischen Modellabweichungen und Transformation

Die einfachste statistische Beschreibung der Modellabweichungen ist ein stochastischer Prozess mit konstanter Varianz. Eine andere, komplexere, Formulierung vergrössert die stochastischen Schwankungen während Niederschlags.

Konstante Varianz: Die einfachste Bias Formulierung ist ein kontinuierlicher Ornstein-Uhlenbeck (OU)-Prozess mit Rückkehr zum Mittelwert (Uhlenbeck and Ornstein, 1930), dessen diskrete Formulierung einem autoregressiven Prozess 1. Ordnung (*AR(1)*) mit normalverteilten unabhängigen Fehlern entspricht. Dieser OU-Prozess führt zu einer Likelihood-Funktion (Gl. (1)) mit Kovarianzmatrix

$$\Sigma_{B_M,i,j}(\psi) = \sigma_{B_{ct}}^2 \exp\left(-\frac{1}{\tau}\left|t_i - t_j\right|\right) \qquad \text{Gl. (3)}$$

, wobei τ die Korrelationszeit ist und σ_{Bct} die asymptotische Standardabweichung der zufälligen Schwankungen um den Mittelwert parametrisiert.

Niederschlagsabhängige Varianz: Eine komplexere Beschreibung koppelt die Varianz des OU-Prozesses quadratisch mit der Niederschlagsintensität x. In analoger Formulierung zu oben ergibt sich für die Kovarianzmatrix der Likelihood-Funktion

$$\Sigma_{B_M,i,j}(\psi, x) = E\left[B_M^2\left(min\left(t_i, t_j\right)\right)\right] \exp\left(-\frac{1}{\tau}\left|t_i - t_j\right|\right) \qquad \text{Gl. (4)}$$

mit

$$E[B_M^2(t_i)] = E[B_M^2(t_{i-1})]\exp\left(-\frac{2}{\tau}\Delta t\right) + \left[\sigma_{B_{ct}}^2 + (\kappa x_{i-\delta})^2\right]\left(1 - \exp\left(-\frac{2}{\tau}\Delta t\right)\right) \qquad \text{Gl. (5)}$$

, wobei κ ein Skalierungsfaktor ist und δ ist ein Zeitversatz, der die Reaktionszeit des Systems auf Niederschlag beschreibt.

Abhängigkeit vom Durchfluss: Die Transformation der Messungen und Modellberechnungen ist gängige Praxis in der Statistik und Hydrologie, um die Zunahme der Varianz mit zunehmenden Messgrössen zu berücksichtigen. Hier untersuchen wir zwei Transformationen zur Stabilisierung der Varianz, die aus unserer Sicht am vielversprechendsten für die Stadtentwässerung sind: die Box-Cox-Transformation und die Log-sinh-Transformation (Wang et al., 2012). Für die Box-Cox Transformation wählten wir λ= 0.35, basierend auf Werten von ähnlichen Studien (Willems, 2012). Details werden in Del Giudice et al. (2013) beschrieben.

2.3 Bayessche Inferenz und Modellprognosen

Um Simulationsmodell mit einer statistischen Beschreibung der Modellabweichungen zu kalibrieren und die Unsicherheit der daraus resultierenden Prognosen zu analysieren, sind vier Schritte notwendig: i) Formulierung von a-priori Verteilungen für die Modellparameter, ii) Kalibration: Schätzung der a-posteriori Verteilungen der Parameter mit Bayesscher Inferenz, iii) probabilistische Vorhersage des Abflusses für die Daten in der Kalibrationsperiode, iv) probabilistische Vorhersage für die Zukunft (Validationsperiode). Die Verteilungen für iii) und iv) werden mit Monte-Carlo-Methoden approximiert. Abschliessend sollte die Qualität der Prognosen beurteilt und die statistischen Annahmen überprüft werden. Eine explorative Analyse der Modellabweichungen kann wertvolle Hinweise zur Verbesserung der Modellstruktur liefern.

Das Prinzip der Bayesschen Parameterschätzung ist, dass Wissen über die Modellparameter θ, formuliert als Wahrscheinlichkeitsverteilung, anhand von Messdaten aufdatiert wird:

$$f(\theta, \psi | y_0, x) = \frac{f(\theta,\psi) f(y_0|\theta,\psi,x)}{\int\int f(\theta',\psi') | f(y_0|\theta',\psi',x) d\theta' d\psi'} \qquad \text{Gl. (6)}$$

Das Integral im Nenner ist zwar nicht analytisch lösbar, jedoch können mit Markov Chain Monte Carlo (MCMC) Algorithmen Realisierungen dieser Verteilung generiert werden. Obwohl dazu mehrere tausend, wenn nicht sogar zehntausend Rechenläufe benötigen werden, ist der Rechenaufwand nicht grösser als der von einfacheren Verfahren (Dotto et al., 2012).

Zur Beurteilung der Modellvorhersagen verwenden wir (i) die Zuverlässigkeit in Bezug auf die Überdeckung der Vorhersageintervalle in der Validationsperiode und (ii) ob die statistischen Annahmen des jeweiligen Fehlermodells für die Kalibrationsperiode erfüllt sind. Zusätzlich berechneten wir die Schärfe der Intervalle und den Nash-Sutcliffe Index.

3 FALLSTUDIE

Wir testeten die Unsicherheit Analyse-Techniken an Daten eines kleinen städtischen Einzugsgebietes in Sadova, Hostivice, in der Nähe von Prag (CZ). Das System hat eine Fläche von 11,2 ha und wird durch ein Trennsystem entwässert. Es ist eine grüne Wohngegend mit einer durchschnittlichen Steigung von circa 2%. Niederschlags- und Abflussmessungen wurden im Sommer 2010 durchgeführt, die zeitliche Auflösung beträgt 2min (Bareš et al., 2010). Die Abflüsse im dem sehr dynamischen System bewegen sich im Bereich von ungefähr 2 l/s bei Trockenwetter bis zu 600 l/s bei starken Regenfällen. Zur Kalibrierung, wählten wir zwei Perioden mit Regenereignissen mit Spitzenintensitäten von 13 bis 65 mm/h.

4 ERGEBNISSE UND DISKUSSION

Wie erwartet, führen die unterschiedlichen Annahmen für die neun Beschreibungen der Modellabweichung zu unterschiedlichen Prognoseintervallen für den Regenwasserabfluss (Abb. 1, links). Im Allgemeinen liefert die explizite Berücksichtigung von Modellabweichungen zuverlässigere Vorhersagen als die mit einem einfachen Fehlermodell und die Prognoseintervalle decken in etwa 95% oder mehr der Validierungsdaten mit dem 95%-Konfidenzniveau ab. Im Gegensatz dazu erzielt das einfache Fehlermodell schlechte Resultate (Abb. 1 links, 1. Reihe). Am schlechtesten ist es in Kombination mit der Log-sinh-Transformation (Abb. 1 links, 1. Reihe, 3. Spalte), wo die 95%-Prognoseintervalle nur knapp 50% der Validierungsdaten enthalten. Das einfache Fehlermodell erfüllt darüber hinaus die statistischen Annahmen weniger, unabhängig von der angewendeten Transformation. Die mit Abstand beste Beschreibung der Modellabweichung in unserer Fallstudie liefert der konstante OU Prozess Log-sinh-Transformation (Abb. 1, links, 2. Reihe, 3. Spalte). Es hat eine hohe Zuverlässigkeit, Schärfe und Nash-Sutcliffe Index.

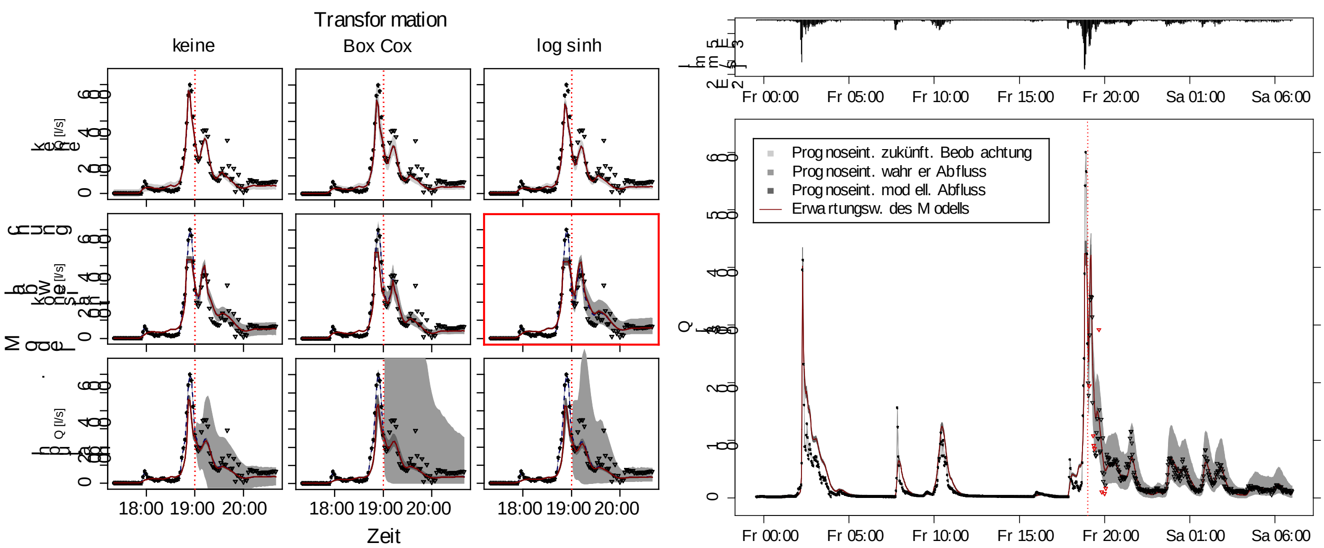

Abbildung 1, Links: Modellprognosen für neun verschiedener Formulierungen der Modellabweichungen; Rechts: Resultate für die beste Beschreibung mit konstantem OU-Prozess und Log-sinh-Transformation (roter Kasten). Man erkennt deutlich wie der OU-Prozess Modellabweichungen in der Kalibrationsperiode korrigiert (links von der vertikalen Linie). Für die Prognose von zukünftigen Abflüssen (rechts von der vertikalen Linie) vergrössert sich die Unsicherheit sichtbar. Ohne Berücksichtigung von Modellabweichungen (links, 1. Reihe) werden die Unsicherheiten der prognostizieren Abflüsse eklatant unterschätzt.

Obwohl die Unsicherheitsintervalle zuverlässiger sind, wenn Modellabweichungen berücksichtigt werden, werden die Erwartungswerte des Kanalnetzmodells leicht schlechter. Dies ist zu erwarten, da die Maximierung der a-posteriori Verteilung in diesem Fall näherungsweise der Minimierung der Summe der Fehlerquadrate entspricht und vergleichsweise die bessere Anpassung liefert. Interessanterweise sind die Vorhersagen mit Niederschlags-abhängiger Formulierung extrem ungenau. Obwohl es konzeptionell überzeugend ist, dass die Vorhersagegenauigkeit bei Regenwetter abnimmt, sind die Prognoseintervalle in unserem Fallbeispiel bei Regenwetter zu wenig scharf und bei Trockenwetter zu eng. Weitere Resultate, sowie eine umfassende Diskussion sind in Del Giudice et al. (2013) beschrieben.

5 SCHLUSSFOLGERUNGEN

Das Ziel dieser Studie war, systematische Modellabweichungen in der Kanalnetzmodellierung zu berücksichtigen. Basierend auf einer statistischen Formulierung mittels eines Gauss'schen Prozess schlagen wir fünf verschiedene Varianten vor, um zu berücksichtigen, dass die Modellprognosen bei Regenwetter ungenauer sind als im Trockenwetterfall. In einer Simulationsstudie für eine Regenwasserkanalisation erzielten wir die besten Ergebnisse mit einer konstanten Varianz und einer Log-sinh-Transformation. Basieren auf den Erfahrungen aus unserer Fallstudie und theoretischen Überlegungen schließen wir, dass:

- Systematische Modellabweichungen in Kanalnetzsimulationen unausweichlich sind, z.B. durch Unsicherheit in den Eingangsdaten, strukturelle Defizite des Modells, und (möglicherweise) systematische Fehler in den hochaufgelösten Durchfluss- oder Wasserstandsmessungen. Deshalb sind probabilistische Prognosen basierend auf einem herkömmlichen unabhängigen Fehlermodell zu optimistisch.
- die einfache Berücksichtigung von Modellabweichungen über einen OU-Prozess mit konstanter Varianz bessere Prognosen liefert, als die niederschlagsabhängige Variante, die konzeptionell überzeugender ist. Der einfache OU-Prozess scheint sich zudem gut für Systeme mit abwechselnden langen Trockenwetterperioden und kurzen intensiven Regenabflüssen zu eignen.
- der vorgeschlagene Ansatz besonders in drei Punkten limitiert ist. Einerseits bedingt die Wahl von a-priori Verteilungen durch den Modellierer ein gewisses Mass an Subjektivität. Andererseits kann diese statistische Beschreibung der systematischen Modellabweichungen "nur" die Auswirkung der verschiedenen Fehlerquellen beschreiben. Zudem sie es so aus, als ob das Fehlermodell jedes Mal an die individuelle Modellierungsstudie angepasst werden muss.

In Zukunft wäre es daher wünschenswert, wenn der Ansatz direkt Informationen über die Ursachen der Abweichungen liefern würde. Eine zusätzliche Herausforderung besteht auch darin, dass für diese Art der Parameterschätzung viele tausend iterative Berechnungen notwendig sind, was für komplexe hydrodynamischen Kanalnetzmodellen zur Zeit noch eine Herausforderung darstellt. In Zukunft könnte das unter Umständen mit stochastischen Ersatzmodellen, z.B. Emulatoren, überwunden werden.

6 REFERENZEN

Bareš, V., D. Stránský, J. Kopecká, and J. Fridrich. 2010. "Monitoring Povodi a Stokove Sito Města Hostivice - Lokalita Sadová. [Monitoring a Sewer Watershed in Hostivice Municipality - Sadová District] (In Czech)". Czech Technical University in Prague.

Del Giudice, D., M. Honti, A. Scheidegger, C. Albert, P. Reichert, and J. Rieckermann. 2013. "Improving Uncertainty Estimation in Urban Hydrological Modeling by Statistically Describing Bias." *Hydrology and Earth System Sciences Discussions* 10 (4) (April 23): 5121–5167. doi:10.5194/hessd-10-5121-2013.

Dotto, Cintia B S, Giorgio Mannina, Manfred Kleidorfer, Luca Vezzaro, Malte Henrichs, David T McCarthy, Gabriele Freni, Wolfgang Rauch, and Ana Deletic. 2012. "Comparison of Different Uncertainty Techniques in Urban Stormwater Quantity and Quality Modelling." *Water Research* 46 (8) (May 15): 2545–2558. doi:10.1016/j.watres.2012.02.009.

Honti, M., C. Stamm, and P. Reichert. 2013. "Integrated Uncertainty Assessment of Discharge Predictions with a Statistical Error Model." *Water Resources Research*: n/a–n/a. doi:10.1002/wrcr.20374.

Reichert, P., and N. Schuwirth. 2012. "Linking Statistical Bias Description to Multiobjective Model Calibration." *Water Resources Research* 48 (9): n/a–n/a. doi:10.1029/2011WR011391.

Uhlenbeck, G.E., and L.S. Ornstein. 1930. "On the Theory of the Brownian Motion." *Physical Review* 36 (5): 823–841.

Wang, QJ, DL Shrestha, DE Robertson, and P. Pokhrel. 2012. "A Log-sinh Transformation for Data Normalization and Variance Stabilization." *Water Resources Research* 48 (5): W05514. doi:10.1029/2011WR010973.

Willems, P. 2012. "Model Uncertainty Analysis by Variance Decomposition." *Physics and Chemistry of the Earth, Parts A/B/C* 42: 21–30. doi:http://dx.doi.org/10.1016/j.pce.2011.07.003.